

- REGRESSION ANALYSIS.

**PEMILIHAN BANDWITDH OPTIMAL PADA ESTIMASI FUNGSI  
REGRESI NONPARAMETRIK**

**SKRIPSI**

MPM 42/04  
Jun  
P



**HENDRA JUNAIDI**

**JURUSAN MATEMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS AIRLANGGA  
SURABAYA  
2004**

**PEMILIHAN BANDWIDTH OPTIMAL PADA ESTIMASI FUNGSI  
REGRESI NONPARAMETRIK**

**S K R I P S I**

**Sebagai Salah Satu Syarat Untuk Memperoleh  
Gelar Sarjana Sains Bidang Matematika  
Pada Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Airlangga**

Oleh :

**HENDRA JUNAIDI**  
NIM. 089711628

**Tanggal Lulus : 29 Juni 2004**

**M I L I E  
PERPUSTAKAAN  
UNIVERSITAS AIRLANGGA  
SURABAYA**

**Disetujui oleh :**

**Pembimbing I,**



**Ir. Elly Ana, M.Si**  
NIM. 131 837 441

**Pembimbing II,**



**Nur Chamidah, S.Si, M.Si**  
NIM. 132 205 653

## LEMBAR PENGESAHAN

Judul : PEMILIHAN BANDWIDTH OPTIMAL  
PADA ESTIMASI FUNGSI REGRESI  
NONPARAMETRIK  
Penyusun : HENDRA JUNAIDI  
NIM : 089711628  
Tanggal Ujian : 29 JUNI 2004

Disetujui oleh :

Pembimbing I,

Pembimbing II,



Ir. Elly Ana, M.Si  
NIM. 131 837 441

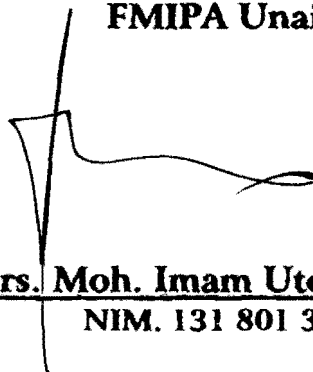


Nur Chamidah, S.Si, M.Si  
NIM. 132 205 653

Mengetahui :

Dekan Fakultas MIPA  
Universitas Airlangga

Ketua Jurusan Matematika  
FMIPA Unair



Drs. H.A. Latief Burhan, M.S.  
NIM. 131 286 709

Drs. Moh. Imam Utoyo, M.Si  
NIM. 131 801 397

Hendra Junaidi 2004, **Pemilihan Bandwidth Optimal Pada Estimasi Fungsi Regresi Nonparametrik**. Skripsi ini dibawah bimbingan Ir. Elly Ana, M.Si dan Nur Chamidah, S.Si, M.Si. Jurusan Matematika FMIPA Universitas Airlangga

## ABSTRAK

Analisis regresi merupakan salah satu alat statistik yang banyak digunakan untuk mengetahui hubungan antara sepasang variabel atau lebih, misalnya Y adalah variabel respon dan X adalah variabel prediktor, maka untuk  $n$  pengamatan, secara umum hubungan variabel itu dapat dinyatakan sebagai berikut :

$Y_i = m(X_i) + \varepsilon_i, i = 1, 2, \dots, n$ , dengan  $m(X_i)$  merupakan fungsi regresi dan  $\varepsilon_i$  adalah error pengamatan yang diasumsikan berdistribusi identik dan independen dengan mean 0 dan varians  $\sigma^2$ .

Tujuan penulisan ini adalah mengestimasi fungsi regresi  $m(X_i)$  dengan pendekatan kernel dan mendapatkan *bandwidth* optimal untuk estimator fungsi regresi. Estimator fungsi

regresi dengan pendekatan kernel digunakan :  $\hat{m}_h(x) = \frac{n^{-1} \sum_{i=1}^n K_h(x - X_i) Y_i}{n^{-1} \sum_{i=1}^n K_h(x - X_i)}$ , dengan  $h$

adalah *bandwidth* dan  $K$  adalah fungsi kernel. Metode untuk memperoleh *bandwidth* optimal adalah metode :

$$CV(h) = n^{-1} \sum_{j=1}^n [Y_j - \hat{m}_h(X_j)]^2 w(X_j), \text{ dengan}$$

$$w(X_j) = K_h(x - X_j) / n^{-1} \sum_{i=1}^n K_h(x - X_i)$$

Sebagai aplikasi digunakan data berat badan terhadap usia dan berat badan terhadap tinggi badan siswa SDN Kali Rungkut I mulai usia 68 sampai 132 bulan untuk anak perempuan. Sehingga didapatkan nilai *bandwidth* optimal pada data berat badan terhadap usia, pada fungsi kernel gaussian adalah 2,70 dan fungsi kernel triangle bernilai 3,00. Sedangkan nilai *bandwidth* optimal pada data berat badan terhadap tinggi badan, pada fungsi kernel gaussian adalah 1,85 dan fungsi kernel triangle bernilai 3,50.

**Kata kunci** : estimasi kernel, bandwidth



Hendra Junaidi 2004, **Optimal Bandwidth Selection in Nonparametric Regression Function Estimation**. This skripsi is under guidance of Ir. Elly Ana, M.Si and Nur Chamidah, S.Si, M.Si. Mathematics Major Subject of Mathematics and Natural Science Faculty. Airlangga University.

## ABSTRACT

Regression analysis is a tool in statistic which usually used to determine relationship between two or more variable, let Y is a response variable and X is predictor variable, then for n observation, the relationship of that variable can be modeled as :

$Y_i = m(X_i) + \varepsilon_i$ ,  $i = 1, 2, \dots, n$ , where the unknown regression function  $m(X_i)$  and observation error  $\varepsilon_i$ , which is assumed distribution identical and independence with mean 0 and variance  $\sigma^2$ .

This research is studied to estimate regression function  $m(X_i)$  with kernel approach and to get optimal *bandwidth* for regression function estimator. Based on kernel

approach, yield regression function estimator :  $\hat{m}_h(x) = \frac{n^{-1} \sum_{i=1}^n K_h(x - X_i) Y_i}{n^{-1} \sum_{i=1}^n K_h(x - X_i)}$ , with h

is *bandwidth* and K is kernel function. Metode for to get optimal *bandwidth* is :

$$CV(h) = n^{-1} \sum_{j=1}^n [Y_j - \hat{m}_h(X_j)]^2 w(X_j), \text{ with } w(X_j) = K_h(x - X_j) / n^{-1} \sum_{i=1}^n K_h(x - X_i)$$

An aplication is used to weigth body data follow age or tall in SDN Kali Rungkut 1 students from age 68 until 132 month for the girl. Than to get value optimal *bandwidth* in data weigth body follow age, in kernel gaussian fanchion is 2,70 and in kernel triangle function is 3,00. Afterwards to get value optimal *bandwidth* in data weigth body follow tall body, in kernel gaussian function is 1,85 and in kernel triangle function is 3,50.

**Key words** : kernel estimation, bandwidth

